ETS GRE®

Assess Ability. Predict Performance.

ETS GRE® Board Research Report
ETS GRE®–17-03
ETS RR–17-06

A Literature Review on Collaborative Problem Solving for College and Workforce Readiness

María Elena Oliveri

René Lawless

Hillary Molloy

December 2017

The report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

****

****

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

GRE-ETS

PO Box 6000

Princeton, NJ 08541-6000

USA

To obtain more information about GRE programs and services, use one of the following:
Phone: 1-866-473-4373
(U.S., U.S. Territories*, and Canada)
1-609-771-7670
(all other locations)
Web site: www.gre.org

*America Samoa, Guam, Puerto Rico, and US Virgin Islands

GRE Board Research Report Series and ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

A Literature Review on Collaborative Problem Solving for College and Workforce Readiness

María Elena Oliveri, René Lawless, & Hillary Molloy

Educational Testing Service, Princeton, NJ

The literature and the employee and workforce surveys rank collaborative problem solving (CPS) among the top 5 most critical skills necessary for success in college and the workforce. This paper provides a review of the literature on CPS and related terms, including a discussion of their definitions, importance to higher education and workforce success, and considerations relevant to their assessment. The goal is to discuss progress on CPS to date and help generate future research on CPS in relation to educational and workforce success.

Keywords Collaborative problem solving; college; higher education; workforce readiness; assessment; noncognitive

doi:10.1002/ets2.12133

Collaborative problem solving (CPS) skills are identified to be important for daily life, work, and schooling in the 21st century. The importance of CPS is best captured in Hutchins's (1995) book, *Cognition in the Wild*, in which the author highlighted the importance of collaboration by asking his readers to scan their immediate environment to identify objects that were not produced collaboratively. Hutchins remarked that in so doing, he was able to uniquely identify only the pebble on his desk. All other objects exemplified teamwork production. Products stemming from collaboration are everywhere. Their production ranges across fields including entertainment, health, nutrition, engineering, and housing, providing impressive evidence of the need for collaboration for progress and development. Yet, as observed by Hesse, Care, Buder, Sassenberg, and Griffin (2015), collaborative skills are neither formally taught nor assessed. Despite the clear need for such skills for workforce readiness, the assessment of 21st-century skills such as CPS lags behind.

In this review paper, we discuss advances related to conceptualization, assessment, and validity considerations related to CPS. We provide a construct model that defines CPS in general and for higher education in particular. We also describe advances in knowledge, skills, and abilities (KSAs) that reveal higher levels of competency in CPS. Finally, we discuss validity evidence (advances and ideas for future research). We aim, through these discussions, to inform current and future efforts in designing, interpreting, and validating CPS assessments.

The goal is to provide an overview of existing CPS definitions, conceptual frameworks or taxonomies, and measurement efforts conducted to date. The second goal is to inform an operational definition of CPS and identify considerations needed for developing a prototype of a CPS assessment. In our review, we cast a wide net for the terminology related to CPS and teamwork. We examined both theoretical and empirical literature, though there is a paucity of the latter.

An overarching definition of CPS was offered by Kyllonen (2012). In his organizational review of the various taxonomies and classifications of diverse 21st-century skills (e.g., Assessment and Teaching of 21st Century Skills [http://www.atc21s.org], Partnership for 21st Century Skills [2012], and National Research Council '[2010]), he defined CPS as "a performance activity requiring groups of students to work together to solve problems" (Kyllonen, 2012, p. 16). Another definition is provided in Hesse et al. (2015), where CPS was described as "a joint activity where dyads or small groups [interact to] execute a number of steps in order to transform a current state into a desired goal state" (p. 39). However, finer-grained definitions are needed if we want to develop tasks that assess the critical elements of CPS. At this more detailed level, we find a lack of coherence with respect to how teamwork and CPS is defined and what behaviors comprise it. CPS is not a well-defined construct, which further complicates efforts to define CPS because researchers typically differ in their definitions of CPS and the elements that comprise it. Researchers have often ignored existing definitions of skills and have not systematically built on previous efforts or definitions of skills provided by other researchers (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995).[1] For example, researchers have used different terminology to describe the

Corresponding author: M. E. Oliveri, E-mail: moliveri@ets.org

same construct. Terms such as *group decision making*, *team cognition, teamwork*, *group work*, *small group problem solving*, *cooperative learning*, *collaborative learning*, and *team collaboration* have all been used interchangeably (O'Neil, Chuang, & Chung, 2004). Moreover, researchers have often used similar labels to refer to different skills.

Literature Review

To gain insights into CPS definitions including its constituent elements, we conducted a multidisciplinary and transorganizational review of higher education and workforce frameworks, spanning fields of education, sciences, business, and management. Occupational literature was also reviewed, especially the literature concerning the military, as much of the research surrounding CPS originated within this particular domain (Cannon-Bowers & Salas, 1997; Dickinson & McIntyre, 1997; Fiore et al., 2010), in which the use of teams to perform complex work functions is pervasive (Shanahan, Best, Finch, & Sutton, 2007). We included higher education and workforce readiness frameworks as the former often serves as preparatory ground for the latter. Each of these various sources presented information at different levels of specificity and detail. For instance, the theoretical literature tended to be descriptive as compared to both the surveys and learning outcomes frameworks, which described CPS skills more globally. We evaluated frameworks and assessments developed for a population of students that were at the end of compulsory education (15 years old and up), spanned the education and workforce domains, and applied to both national and international contexts.

There is voluminous literature on CPS spanning decades of research conducted by both individual researchers and organizations. Stevens and Campion (1994), who conducted early research in identifying specific KSAs associated with teamwork, provided one of the earliest frameworks on CPS. In their framework, they outlined the KSAs to recognize the obstacles to collaborative group problem solving and to utilize the proper degree and type of participation, in contrast to the identification of global personality traits (e.g., helpfulness, supportiveness) as researchers had done previously (p. 505). Additionally, the authors focused on the individual's role in successful team performances rather than focusing on the group or the team as the unit of analysis, which was critical to advance the development of an assessment that could provide information about individual test takers. Finally, the authors identified transportable teamwork competencies (conflict resolution, CPS, communication, goal setting and performance management, and planning and task coordination) believed to be generalizable across contexts.

The framework by Cannon-Bowers et al. (1995) expanded this work and made two important contributions. First, the authors expanded the set of five transportable competencies suggested by Stevens and Campion (1994) to eight, with the argument that the additional competencies are also central to CPS. The eight competencies include (a) adaptability, (b) shared understanding of the situation, (c) performance monitoring and feedback, (d) leadership, (e) interpersonal relations, (f) coordination, (g) communication, and (h) decision making. Cannon-Bowers et al. further explicated distinctions between team-generic and team-specific competencies, as well as task-generic and task-specific competencies. Team-generic competencies are those that an individual possesses that can be used in any group work, regardless of its composition, such as communication skills, leadership skills, and attitudes toward teamwork. Team-specific competencies, conversely, are those tied to the composition of the group, such as "knowledge of teammates' characteristics, specific compensation strategies, and team cohesion" (Cannon-Bowers et al., 1995, p. 338). Similarly, task-specific competencies are KSAs that are relevant to a specific task (e.g., an understanding of the specific roles and responsibilities of each teammate), whereas task-generic competencies are those that an individual uses across a variety of tasks (e.g., general planning skills). Competencies deemed "transportable" are those that are both task-generic and team-generic. Since we are looking to develop an assessment that can predict a test taker's ability to collaborate with a wide variety of teams across an array of contexts, transportable competencies are the ones that we aim to target in the development of a CPS measure for higher education.

Several research teams further expanded the abovementioned frameworks and elaborated on the description of the skills they comprise. For example, the notion of a *shared mental model* as being central to team performance was included (Aguado, Rico, Sánchez-Manzanares, & Salas, 2014; Cooke, Salas, Cannon-Bowers, & Stout, 2000; Klimoski & Mohammed, 1994; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Processes related to shared mental models are thought to enable team members to have common ways of interpreting cues, allocating resources, and making compatible decisions (Cooke et al., 2000). Several scholars (Aguado et al., 2014; Cannon-Bowers et al., 1995; Zaccaro, Rittman, & Marks, 2001) reiterated the importance of including *team leadership* in a conceptual framework. Zaccaro et al. described how it is composed of four superordinate components: (a) information search and structuring, (b) information

use in problem solving, (c) managing personnel resources, and (d) managing material resources. Recognizing that CPS is a dynamic process, Marks, Mathieu, and Zaccaro (2001) suggested including a *temporal aspect* in teamwork and collaboration and indicated that collaborative actions occur along a series of simultaneous tasks, including creating team goals and strategies, monitoring and coordinating the actions of teammates, and managing tasks occurring over time. Similar to Stevens and Campion (1994), M. M. Chiu's (2000) taxonomy explained collaboration and group problem-solving processes in detail, providing a focus not only on social interactions but also on the individual's role within the team. Chiu's research also provided exemplars of directly measurable behaviors and hence of importance to assessment development. Examples of the kinds of social interactions that students engage in when problem solving include (a) taking turns and working cooperatively; (b) providing tangible demonstrations of solutions to problems; and (c) listening attentively, being responsive to each other's' input, and connecting ideas with each other through full participation.

Beyond the aforementioned taxonomies developed by individual researchers, we also reviewed taxonomies developed by organizations that were developed for the assessment of CPS within a large-scale assessment context. These included higher education and workforce frameworks. Tables 1 and 2 provide a summary of the higher education and workforce frameworks we reviewed, respectively. One example is the framework developed by the Programme for International Student Assessment (PISA) CPS assessment commissioned by the Organisation for Economic Co-operation and Development (OECD). The OECD (2013b) proposed that CPS comprises three competencies: (a) establishing and maintaining shared understanding, (b) taking appropriate action to solve a problem, and (c) establishing and maintaining team organization. Table 3 presents frameworks that span the higher education and workforce contexts.

The goal of the review of the frameworks summarized in Tables 1 – 3 and the empirical research on CPS was to develop an operational definition of CPS that would be applicable to the higher education and workforce contexts and that would contain overlapping KSAs across contexts. We used four sources of evidence. First, we used an iterative approach that consisted of reviewing higher education student learning outcomes literature and noted the skills and behaviors that required working with groups in various capacities (e.g., collaboration, group communications, or group projects). We reviewed journal articles, book chapters, and books and identified the various skills and behaviors that were suggested to play a role in successful team interactions while conducting various activities or projects. Our review of the literature yielded extensive lists of skills and behaviors that we summarize below.

Second, this goal of achieving an operational definition was informed by consultation with experts, including research scientists, test developers with expertise in international populations, and psychometricians. Our collaborators brought expertise in evidence-centered design, design spaces, assessment strategies, prototyping, the development of CPS items for the PISA, and the development of items measuring noncognitive constructs. They also had expertise with various item types used to assess noncognitive constructs (e.g., forced-choice items, ipsative, and quasi-ipsative measures) in the context of CPS with human and virtual agents. Together, we engaged in an iterative process where five experts reviewed various versions of the proposed model and gave feedback, which led to at least three revision cycles with modifications and updates at each stage.

Third, we attended the Innovative Assessment of Collaboration 2014 conference, which brought together researchers and international experts in organizational teaming, educational collaboration, tutoring, simulation, gaming, and statistical and psychometric process modeling for insight into the development of reliable and valid collaborative assessments. The meetings focused on team performances in organizations, simulated teamwork environments, and CPS in educational settings (ETS, 2014b).

The final source of evidence we relied on was the Council for the Advancement of Standards in Higher Education (CAS; Strayhorn, 2006), which was developed to guide higher education professionals in the assessment of 16 student learning and development outcomes, such as leadership development and effective communication. As summarized in Table 1, the CAS standards provide definitions of CPS, among other learning outcomes; describe the theoretical underpinnings of the constructs; provide a list of related variables; and give suggestions for how to measure these outcomes. CPS was named as one of the critical learning outcomes in the CAS.

An Operational Definition of CPS

Figure 1 illustrates our proposed definition of CPS. We suggest that CPS comprises four components: teamwork, communication, leadership, and problem solving. Each has multiple associated skills. Gradations (examples of KSAs that can be demonstrated at various proficiency levels for some of the constituent elements such as teamwork and communication)

Table 1 Higher Education Frameworks

Framework	Who provided input	What was done	Method	Intended outcomes
The Council for the Advancement of Standards in Higher Education (CAS)	Delegates from 36 higher education associations, student affairs offices, and nonprofits in United States and Canada	Developed professional practice standards for undergraduate and master's level students; provided outcomes assessment tools to educators wanting to assess student learning development	NS	Help guide higher education administrators and institutions that work with college students and master's-level students to promote the assessment of students' growth, goals, and behaviors
Framework for 21st Century Learning: Communication/ Collaboration, Critical Thinking, & Problem Solving	Teachers, education experts, business leaders	Defined the KSAs students need in work and life; and the support systems that will enable this	NS	Help practitioners integrate these skills into core academic subjects
The Lumina Foundation's Degree Qualifications Profile	Four original authors, expert reviewers, faculty	Developed reference points for what students should know and be able to do for associate's, bachelor's, and master's degrees	NS	Inform students on knowledge and abilities required for earning undergraduate or master's degrees
Liberal Education and America's Promise (LEAP)	Educators from hundreds of colleges and universities, employers	Described the kinds of learning that students need from college and how to help them achieve it	Expert opinions, business reports, accreditation requirements	Educate students and the public about quality liberal education outcomes and advocate their achievement by college students
Tuning Project	Original authors, over 100+ universities	Defined commonly accepted professional and learning outcomes and generic & subject-specific competencies	Discussions; benchmark papers; questionnaires; online websites	Create transparency in educational structure and promote further innovation through communication and identification of good practices
The Assessment and Teaching of 21st Century Skills (ATC21S)	250 researchers from around the world	Defined 21st century skills (communication, collaboration, problem-solving, digital and ICT literacy); developed ways to transform the teaching, learning, and assessment of these skills	White papers from each working group	Help prepare entry-level workers with practical skills to "create, build, and help sustain information-rich business" by providing education systems with curricular recommendations, innovative assessments, and teaching–learning resources

Note. NS = not specified.

Table 2 Workforce Frameworks

Framework	Who provided input	What was done	Method	Intended outcomes
O*NET Competencies Framework	Employees, job analysts	Described the KSAs, interests, and general work activities associated with each occupation	Survey, in-depth interviews	Public knowledge and dissemination
Workforce Readiness Project	431 employers, 12 HR executives	Determined the corporate perspective on how ready entry-level workers in the United States are, by level of educational attainment	Survey, in-depth interviews with 12 HR executives	To encourage the business community, educators, policy makers, and students to augment workforce readiness preparation by providing a definition of workforce readiness
Hewlett Foundation's 21st Century Competencies	A wide range of literature was summarized, including literature on human capital and work readiness	Determined which worker competencies are most important for the 21st century and how possessing them makes a difference in educational and economic outcomes for individuals and employers	Literature review	Promote the development of programs and systems that will assist in acquiring and assessing skills that are critical to the economy
ETA Competency Model Clearinghouse's General Competency Model Framework	Industry associations, labor organizations, educators and I/O psychology, and subject matter experts	Developed a model describing competencies on nine tiers, hierarchically organized in three tiers: foundational, industry- and occupational-related competencies	NS	Inform and support the education and training of a competitive workforce by determining the necessary requirements of workers based on business and employer needs
DOL-ETA SCANS Report	Job experts, occupational analysts	Determined the major types of skills required to enter employment, demonstrated the level of importance of the skills in sample jobs, identified specific job tasks that illustrated the use of the skills	Expert opinions, literature review, analysis of 50 jobs, structured interviews	Help make high school courses contain information relevant to the workforce; help employers to make sure their workers have relevant skills

Note. NS = not specified.

Table 3 Combined Higher Education and Workforce Frameworks

Framework	Who provided input	What was done	Method	Intended outcomes
Council of Graduate Schools	University leaders, business leaders from nonprofit and for-profit sectors, Council of Graduate Schools, ETS	Described the link between graduate education and the workforce, described the professional skills that employers seek and currently do not find in entry-level graduate students, and made a call to strengthen the link	Expert input, interviews with employers, survey of graduate deans, survey of graduate students	Help develop professional skills relevant to students' careers in graduate school by outlining roles of stakeholders such as universities, employers, and policy holders and provide recommendations to each of these groups on how they can clarify career pathways for graduate students.
AAC&U/HART Research Associates	400 employers with 25+ employees; 613 college students	Determined employers' and college students' perspectives on (a) which learning outcomes are most important to succeed in today's economy, (b) how prepared college graduates are, and (c) importance of project-based learning in college	Online surveys	NS
National Research Council's 21st Century Skills	Council committee members	Defined and summarized the literature on 21st century skills and (a) how these skills relate to traditional academic skills, (b) importance of these skills to educational and work success, and (c) how these skills can be taught, assessed, and learned	Research literature review from several disciplines, NRC workshops, other key papers	Educate public on the research on teaching/learning of key skills needed in the 21st century

Note. NS = not specified.

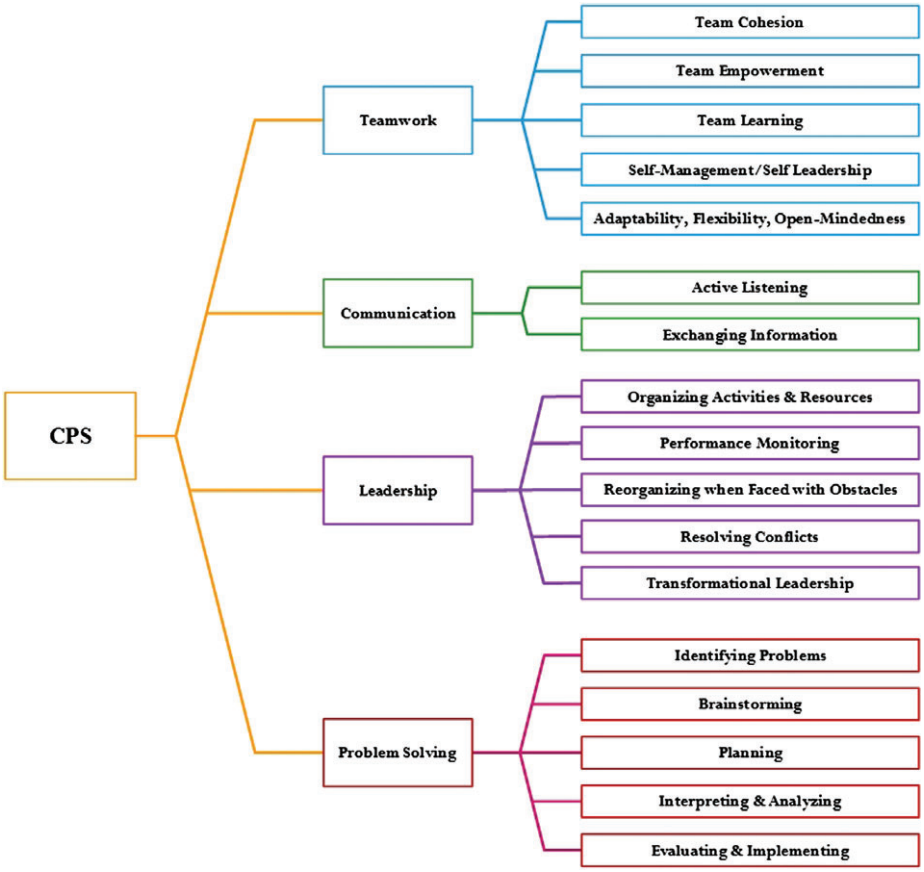

Figure 1 The specific skills associated with each of the four components of our proposed higher education readiness CPS taxonomy.

contextualized in higher education have been recently developed by Lench, Fukuda, and Anderson (2015). The existing gradations inform us of what types of behaviors are associated with each component for various levels of the components. The authors suggested that individuals at every age level may fall along different stages of the developmental continuum depending on life experiences and learning paths. In the report, the authors also provided specific ways in which various skills map onto specific higher education activities and processes. Similarly, we also map higher education processes to specific behaviors in our operational definition to elucidate the connection between the behaviors in our proposed definition and the higher education and workforce preparatory activities students are likely to encounter.

A Description of the Proposed Constituent Elements

In what follows, we describe each component of our proposed definition. As mentioned, each component is aligned with the definitions from and connections to existing higher education student learning outcomes frameworks and existing assessments of CPS, particularly those that focus on transportable competencies.

Teamwork

Teamwork consists of five main skills. These include processes related to promoting (a) team cohesion, (b) team empowerment, (c) team learning, (d) self-management and self-leadership, and (e) attitudes of open-mindedness, adaptability, and flexibility.

Team Cohesion

Team cohesion involves having increased knowledge of teammates and the activities they find enjoyable, and it is an important aspect of team performance (Salisbury, Carte, & Chidambaram, 2006). Although its transportability is debated

by different frameworks (Aguado et al., 2014), certain team cohesion behaviors are relevant to the higher education and workforce contexts, such as encouraging team members' contributions in meetings and understanding group dynamics while carrying out group projects (Association of American Colleges and Universities [AAC&U], 2011; Strayhorn, 2006).

Team Empowerment

Team empowerment refers to the ability to be committed to one's team and the ability to empower team members by challenging their opinions and motivating them to take on additional challenges and interact with others to maximize their strengths (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Dickinson & McIntyre, 1997; Marks et al., 2001). In higher education and the workforce, team empowerment includes motivating and inspiring action in others, expressing confidence in the assigned task and the team's ability to accomplish a goal, contributing to the achievement of a common goal of the group, and leveraging the strengths of others to achieve common goals (AAC&U, 2011; Casner-Lotto & Barrington, 2006; González & Wagenaar, 2003; Markle, Brenneman, Jackson, Burrus, & Robbins, 2013).

Team Learning

Kukenberger, Mathieu, and Ruddy (2015) defined team learning as the extent to which one's knowledge and ability have increased as the result of being a team member. Kostopoulos, Spanos, and Prastacos (2013, p. 1453) described it as a "collective property that although it builds on, cannot be reduced to, individual contributions." Evidence of team learning might include reflective statements provided by students at the end of a group project when they are asked to summarize their contribution to the completion of the team assignment.

Self-Management and Self-Leadership

Rousseau and Aubé (2010) defined team self-managing behaviors as those in which team members collectively distribute tasks. Each person is responsible for taking care of the planning and execution of their assigned tasks to meet the goals of the team (Burrus, Jackson, Xi, & Steinberg, 2013; Ennis, 2008; Markle et al., 2013). Team members are responsible for monitoring their own performances during this process. They are also responsible for making any adjustments to their plans if obstacles are encountered. Self-leadership involves the monitoring of one's performance on a task as well as demonstrating one's influence on the team goals to achieve success (AAC&U, 2011; González & Wagenaar, 2003). An example of a behavior associated with self-management and self-leadership is demonstrating the monitoring of one's performance—for example, by looking for one's own improvements to increase personal or group efficacy (AAC&U, 2011; Burrus et al., 2013). In the context of higher education and the workforce, manifestations of these skills may involve ensuring that one's assignments are all completed on time in a thorough and comprehensive way to advance a team project (AAC&U, 2011; Adelman, Ewell, Gaston, & Schneider, 2014).

Open-Mindedness, Adaptability, and Flexibility

A vital part of teamwork is the ability to demonstrate the skills of being open-minded, adaptable, and flexible in order to enable the best performance of the team. Behaviors that fall within these three skills include being open-minded in order to incorporate others' ideas and feedback into the group's project; responsiveness to diverse perspectives, ideas, and values; the ability to work for and accept necessary compromises as a team member; able to adapt one's behavior to increase its suitability for others; flexibility, which includes the ability to change one's behavior or thinking when conditions change; and the ability to accept ambiguity (Dickinson & McIntyre, 1997; Hesse et al., 2015; OECD, 2013b). Across higher education and workforce contexts, open-mindedness and adaptability were viewed to matter in relation to being open and responsive to diverse perspectives, ideas, and values; seeking and providing constructive feedback; and identifying the rationale behind teammates' opinions and perspectives (AAC&U, 2011; Markle et al., 2013). The expression of flexibility involves the willingness to make necessary compromises to accomplish a common goal (Casner-Lotto & Barrington, 2006; Ennis, 2008).

Communication

A second constituent element of CPS is communication. Two skills are central to communication: (a) active listening and (b) information exchange.

Active Listening

Active listening serves to ensure that a message is received as stated (Stevens and Campion, 1994). It involves the ability to interpret others' nonverbal cues, to listen nonjudgmentally, to probe the speaker for clarifying information, and to give others the opportunity to speak (Cannon-Bowers et al., 1995; Dickinson & McIntyre, 1997; Fiore et al., 2010). It also involves giving interlocutors full attention as they speak, making a conscious effort to understand the speaker's points, posing appropriate follow-up questions, and refraining from inappropriate interruptions. The behaviors were identified as important in both higher education and workforce frameworks (Burrus et al., 2013; Strayhorn, 2006). Moreover, as a testament to the value of active listening skills, Woolley, Chabris, Pentland, Hashmi, and Malone (2010) found that conversational turn-taking was important to team success, which is to say that teams performed better across a variety of tasks when team members were given equal opportunities to speak than when a few select individuals dominated the conversations.

Exchanging Information

Exchanging information includes the ability to send congruent messages that are clear, accurate, and validating. The exchanged messages are transactional rather than personal, thus making this skill distinct from the communicative aspect of team cohesion. The competent individual is able to provide relevant project updates to teammates (Fiore et al., 2010; Marks et al., 2001; OECD, 2013b). During the exchange of information, individuals take turns to ensure that team members are given the opportunity to speak. This kind of interaction not only includes prompting and responding to what others say, but also helps ensure that their nonverbal cues agree with their verbal cues.

The need to exchange information was mentioned as a central aspect of CPS across workforce and higher education frameworks. González and Wagenaar (2003) stated that it is necessary to communicate with nonexperts in one's field as well as to effectively communicate with all team members to achieve the goals and objectives of the team. For example, exchanging information is necessary to keep parties informed of the progress and any changes to a given project to meet project timelines (Ennis, 2008). Strayhorn (2006) suggested that in higher education it is needed to articulate abstract ideas effectively. The frameworks by the International Society for Technology in Education (ISTE, n.d.) and the Partnership for 21st Century Skills (2012) stated that students should be able to communicate ideas and information effectively and articulate thoughts and ideas across multiple audiences using diverse media and formats.

Leadership

We suggest that there are five skills that are relevant to leadership within the context of higher education and workforce readiness: (a) organizing activities and resources, (b) monitoring performances, (c) reorganizing when faced with obstacles, (d) resolving conflicts, and (e) demonstrating transformational leadership.

Organizing Activities and Resources and Performance Monitoring

Kozlowski and Ilgen (2006) indicated that organizing activities and resources and performance monitoring are central to setting and meeting realistic team learning goals. An effective project manager will be able to balance the need to advance project goals key to the group's success with the need to respect time and resource constraints (Ennis, 2008; Hesse et al., 2015; O'Neil, Chung, & Brown, 1997). This involves communicating clear roles, responsibilities, and expectations to team members and distinguishing between problems that can be solved either individually or within a team (Dickinson & McIntyre, 1997; OECD, 2013b; Stevens & Campion, 1994). Organizing activities involves the ability to articulate a vision for the project that can be communicated to and shared by the team.

Key behaviors include clearly and effectively defining the roles and responsibilities of team members to ensure that team goals are understood, the information is shared, and team members have the necessary resources to perform their assigned tasks (C.-J. Chiu et al., 2011; OECD, 2013b; Marks et al., 2001; O'Neil et al., 1997; Stevens & Campion, 1994). As suggested by the higher education and workforce frameworks, these activities involve the ability to (a) design and manage projects (González & Wagenaar, 2003); (b) plan and schedule tasks to complete assigned work on time (Ennis, 2008); (c) allocate time and resources effectively, including coordinating efforts with all parties involved (Ennis, 2008);

(d) evaluate and maintain the quality of work (González & Wagenaar, 2003); and (e) keep track of details to ensure work is performed accurately and completely (Ennis, 2008).

Reacting to Obstacles and Resolving Conflicts

The ability to resolve conflicts involves reacting to or anticipating obstacles preemptively (AAC&U, 2011; Hesse et al., 2015; Marks et al., 2001; O'Neil et al., 1997). It also involves identifying and correcting gaps, errors, or misunderstandings that arise during the execution of tasks (OECD, 2013b). Ideally, conflict is confronted and win – win strategies are used to resolve conflict directly and constructively. Instead of focusing on personal gain, it may be useful to encourage solutions that benefit all team members to address common goals (AAC&U, 2011; Stevens & Campion, 1994). Alternatively, conflict can be resolved by incorporating the needs and viewpoints of all parties. In either case, insights regarding opposing or different views should be sought to hear all members' views and minimize misunderstandings (Marks et al., 2001).

In higher education, the Lumina DQP (Adelman et al., 2014) highlighted the importance of negotiating successful strategies when conducting group research. The AAC&U (2011) provided examples of such strategies. These include addressing destructive conflict directly and constructively to help manage or resolve potential conflicts to strengthen the team and the future effectiveness of the team and project. In the workforce, employers want employees who can anticipate obstacles to project completion, develop contingency plans to address the obstacles, and take corrective action when projects go off track. Desirable employees are those who can bring others together to reconcile differences, handle conflicts maturely through a mutual give-and-take approach, and actively promote mutual goals and interests (Ennis, 2008).

Transformational Leadership

Transformational leadership describes the ability of a team member to motivate other team members to work hard to attain or exceed goals and take on more difficult or challenging tasks. Such a person brings added value to the team. Mitchell et al. (2014) argued that positive team dynamics, a receptiveness to diversity, and positive motivation are all the results of effective transformational leadership.

Problem Solving

Problem solving focuses on the creation of strategies to answer a given problem, dilemma, or open-ended question. It is defined as "the process of designing, evaluating, and implementing a strategy to answer an open-ended question or achieve a desired goal," wherein the solution or the strategy required to arrive at a solution is not readily apparent (AAC&U, 2011 p. 1; Murray, Owen, & McGaw, 2005; OECD, 2013a). Under the PISA 2012 problem-solving framework, this process involves four overarching stages: (a) exploring and understanding, (b) representing and formulating, (c) planning and executing, and (d) monitoring and reflecting (OECD, 2013a). Conversely, the problem-solving VALUE rubric, authored by the AAC&U (2011), includes six dimensions: (a) define the problem, (b) identify the strategies, (c) propose solutions and hypotheses, (d) evaluate potential solutions, (e) implement the solutions, and (f) evaluate the outcomes. Five stages of problem solving are commonly identified in the literature: (a) identifying and defining a problem, (b) brainstorming, (c) planning, (d) interpreting and analyzing information, and (e) evaluating and implementing solutions. Although problem solving is important for higher education and workforce readiness, in alignment with CPS, we focus on the interactions of individuals during the various problem-solving stages rather than the capability of each individual to problem solve, which is already integrated into existing higher education assessments.

Collaborative Problem Solving Assessments

Our second objective is to provide exemplars of existing CPS assessments. Table 4 illustrates a selection of existing teamwork and CPS assessments spanning fields including business, education, health and medicine, and the military. Examples include the VIEW assessment (Creative Problem Solving Group, 2013), the Teamwork-KSA Test based on the taxonomy by Stevens and Campion (1994, 1999), and the CCSSO Workplace Readiness Assessment Consortium (Grummon, 1997). These assessments measure CPS and related constructs using multiple themes, some of which overlap. For example,

teamwork, communication, problem solving, and leadership were found frequently across assessments. Some dimensions or themes were found less frequently, such as situational awareness, workload management, and product quality.

Assessment Format

Table 4 lists the task types that have been used to measure CPS and related skills and includes self-assessments that use Likert scales or forced-choice options, situational judgment tests (SJTs), third-party evaluations, and observation tools (e.g., behavioral checklists, audio- and videotaped observations), and the analysis of think-aloud protocols (e.g., Brannick, Prince, Prince, & Salas, 1993; Oser, McCallum, Salas, & Morgan, 1989). We elaborate on the advantages and disadvantages of various item formats later in the paper.

General Versus Domain Specific

Teamwork and CPS assessments administered in the health and medical fields are often domain specific, which is to say that they aim to assess CPS skills in highly specific contexts that require some prerequisite content knowledge. For example, the University of Texas Behavioral Marker Audit Form (Thomas, Sexton, & Helmreich, 2004) is used to assess teamwork skills during simulations of neonatal resuscitation. Another example is the Anaesthetists' Non-Technical Skills test (ANTS; Flin, Glavin, Maran, & Patey, 2012), which also takes the form of a high-fidelity simulation to assess anesthetists' CPS skills in situations closely resembling those they may encounter in the workplace.

On the other hand, VIEW (Creative Problem Solving Group, 2013), the Teamwork Competency Test (TWCT; Aguado et al., 2014), and the Teamwork-KSA Test (Stevens & Campion, 1994, 1999) were developed to assess general team competencies (i.e., transportable competencies). Similarly, PISA was developed to "use problem situations and contexts relevant to 15-year-old students that tap generalized problem-solving skills, but do not rely on specialized knowledge" (OECD, 2013b, p. 14). Transportable skills are of particular interest to a higher education assessment context, as test takers may be submitting their scores to graduate programs across a wide array of academic and professional disciplines and in support of the teaching and learning of skills that are relevant when transitioning from higher education to the workforce.

Test and Scale Reliability

Some CPS assessments contain and provide scores for more than one subscale. For example, VIEW has three subscales: orientation to change, manner of processing, and ways of deciding. Each subscale has reliabilities above .80 (Treffinger et al., 2014). An advantage of using subscale scores is the ability to augment the information obtained on each dimension(s). A disadvantage is that subscale scores typically have lower reliability. For example, Athanasaw, 2003 (as cited in Aguado et al., 2014) obtained a coefficient of .66 for the complete teamwork-KSA test scale and reliabilities ranging between .25 and .48 for each of the five factors (subscales).

Validity of Existing CPS Assessments

Our review of validity evidence for existing assessments of CPS suggested that the number and quality of studies varied widely. It ranged from evidence on the number of dimensions extractable from the measures, whether the measures predicted and augmented aptitude and cognitive measures typically utilized in assessing team performance. Moreover, most studies were conducted in the workforce within specific contexts as compared to the use of teamwork measures in the higher education context.

Construct Validity

Few construct validity studies on CPS measures are available. Among existing studies, two situations typically arose: (a) the empirical model did not support the hypothesized structure, and (b) a suggestion was made to revise the existing scale and assess fewer factors than originally hypothesized. For example, a study conducted by Oliveri, McCaffrey, Holtzman, & Ezzo (2014) yielded improved fit for a two-factor model rather than the originally hypothesized six-factor structure associated with the *ETS*® Personal Potential Index (PPI). The authors suggested assessing fewer factors in a revised scale.

Table 4 Assessments of Collaborative Problem Solving (CPS) and Related Constructs

Assessment and developer name	Field	Description of the assessment	Delivery mode	Themes and topics	Inferential level
Self-assessments are of two kinds: (a) In the context of Likert-type self-assessments, respondents are asked about their personal preferences, beliefs, tendencies, and behaviors; and (b) in the context of forced-choice assessments, test takers are presented with multiple statements linked to different dimensions and are asked to select the statement that is closest to the way they perceive themselves.					
VIEW: An Assessment of Problem Solving Style (Creative Problem Solving Group, 2013)	Business and education	Contains 34 Likert scale forced-choice hybrid items to create a profile of the test taker's problem-solving style and preferences	Web-based	• Orientation to change • Manner of processing • Ways of deciding	Individual
Tailored Adaptive Personality Assessment System (TAPAS) (Drasgow et al., 2012)	Military	Contains forced-choice items to assess certain noncognitive characteristics deemed valuable for entry-level soldier performance and retention. These include subcomponents of the Big Five personality traits	Web-based	• Extroversion: dominance, sociability, excitement seeking • Agreeableness: warmth, generosity, co-operation • Conscientiousness: industriousness, order, self-control, responsibility, virtue, traditionalism, physical condition • Emotional stability: adjusted, even-tempered, happy • Openness: intellectual, ingenuity, tolerance, curiosity, depth, aesthetics	Individual
ETS *WorkFORCE*® Assessment for Job Fit (ETS, 2014a, 2014b)	Education and workforce	Computerized adaptive self-assessment with approximately 100 forced-choice items that measure six key elements of workplace success. It is based on the U.S. Army's Tailored Adaptive Personality Assessment (TAPAS)	Web-based	• Initiative and perseverance • Responsibility • Teamwork and citizenship • Customer service orientation • Problem solving and ingenuity • Flexibility and resilience	Individual

Table 4 Continued

Assessment and developer name	Field	Description of the assessment	Delivery mode	Themes and topics	Inferential level
Teamwork Competency Test (TWCT; Aguado et al., 2014)	Education and workforce	Targets the 14 subcomponents of teamwork as outlined by the Stevens and Campion (1999) framework. It contains "36 items in a 4-point frequency scale format and drafted in 'observable behaviors' statements" (Aguado et al., 2014, p. 116)	Paper-and-pencil	• Conflict resolution • Collaborative problem solving • Communication • Goal setting and performance management • Planning and task coordination	Individual
Situational judgment tests (SJTs) are a type of scenario-based assessment. Test takers are presented with a particular situation by text (low fidelity) or video (high fidelity) and are asked to determine how someone should react under the presented circumstances. After being presented with the scenario, they are given a set of possible reactions and asked to either select one option or to rate the appropriateness of each option.					
The Teamwork-KSA Test (Stevens & Campion, 1994, 1999)	Higher education and workforce readiness	Contains 35 multiple-response items. Students are asked to respond to how they would act across different situations potentially arising within a work team	Paper-and-pencil	• Conflict resolution • Collaborative problem solving • Communication • Goal setting and performance management • Planning and task coordination	Individual
Predictor Situational Judgment Test (PSJT; Waugh & Russell, 2005)	Military	Contains written scenarios with 20 items designed to improve personnel selection and classification for the U.S. Army. Assesses judgment and decision-making skills across situations commonly encountered during the first year in the Army (Knapp & Heffner, 2009)	Paper-and-pencil	• Adaptability to changing conditions • Relating to and supporting peers • Teamwork • Self-management • Self-directed learning	Individual
In third-party evaluations, a third-party rates test takers on behaviors such as knowledge and creativity.					
ETS® Personal Potential Index (PPI; ETS, 2009)	Graduate school readiness	A web-based evaluation system containing 24 Likert-type items distributed to up to five evaluators of the applicant's choosing. It is designed to supplement the Graduate Record Examination and provide a more complete picture of an applicant's graduate school readiness	Web-based	• Knowledge and creativity • Communication • Teamwork • Resilience • Planning and organization • Ethics and integrity	Individual

Table 4 Continued

Assessment and developer name	Field	Description of the assessment	Delivery mode	Themes and topics	Inferential level
Observational tools (i.e., behavioral marker systems; behavioral checklists) are used while test takers are engaged in either a real or simulated performance task to note whether (un)desirable behaviors are displayed during a given task. Raters indicate (a) whether or not a particular behavior was observed, (b) its quality, and (c) its frequency. They are typically developed based on the analysis of performances that contributed to (un)successful outcomes (Daimler-Und & Benz-Stiftung, 2001).					
University of Texas Behavioral Marker Audit Form (Thomas et al., 2004)	Health and medicine	Used by trained raters while observing simulations of neonatal resuscitation (M. A. Rosen et al., 2010)	High-fidelity simulation	• Information sharing • Inquiry • Assertion • Sharing intentions • Teaching • Evaluation of plans • Workload management • Vigilance/environmental awareness • Overall teamwork • Leadership	Group
Performance Assessment of Communication and Teamwork (PACT; C.-J. Chiu et al., 2011)	Health and medicine	Contains three observation tools: (a) novice observer form, (b) long form for experienced observers, and (c) video coding sheet to assess teamwork and communication during simulated scenarios, including capturing the frequency and quality of desired teamwork and communication behaviors	High-fidelity simulation	• Team structure • Leadership • Situation monitoring • Mutual support • Communication	Group
Anesthetists' Non-Technical Skills (ANTS; Flin et al., 2012). Developed by the University of Aberdeen Industrial Psychology Research Center and the Scottish Clinical Simulation Centre	Health and medicine	Measures observable nontechnical skills associated with good practice in task management, team working, situational awareness, and decision making	High-fidelity simulation	• Task management • Team working • Situation awareness • Decision making	Individual

Table 4 Continued

Assessment and developer name	Field	Description of the assessment	Delivery mode	Themes and topics	Inferential level
Communication and Teamwork Skills (CATS; Frankel, Gardner, Maynard, & Kelly, 2007)	Health and medicine	Observer protocol designed to assess the examinee's communication and teamwork skills in clinical settings. The observer uses the form to capture the frequency and quality of desirable behaviors	High-fidelity simulation	• Coordination • Cooperation • Situational awareness • Communication	Group
Observational Teamwork Assessment of Surgery (OTAS; M. A. Rosen et al., 2010)	Health and medicine	Measures task work and teamwork surrounding the surgical processes to assess communication, leadership, monitoring, and cooperation. Rated by an expert using a 7-point Likert scale to score behaviors	High-fidelity simulation	• Communication • Leadership • Coordination • Monitoring • Cooperation	Group
Team Problem Solving Assessment Tool (TPSAT) (Rotondi, 1999)	Health and medicine	Designed to "aid managers and other practitioners convene and coach teams more effectively" (Rotondi, 1999, p. 205). For each variable, the evaluator assigns a score between 0 and 100	High-fidelity simulation	• Customer's values • Team member expertise • Team interaction style • Systematic problem exploration and solution development • Meeting facilitation • Pressure to solve the problems • Problem definition • Team member participation • Written logs of meetings	Group

Table 4 Continued

Assessment and developer name	Field	Description of the assessment	Delivery mode	Themes and topics	Inferential level
Anti-Air Teamwork Observation Measure (ATOM; Smith-Jentsch, Johnston, & Payne, 1998)	Military	Uses a three-step process consisting of a four-dimensional model of teamwork in the form of qualitative notes and ratings	High-fidelity simulation	• Information exchange • Communication • Supporting behavior • Leadership and initiative	Group
Blended assessment consisting of a prototype observation tool, third-party evaluation of the final product, and a self-assessment					
CCSSO Workplace Readiness Assessment Consortium (Grummon, 1997)	Secondary education	Designed to help instructors give students feedback on their teamwork skills and to help instructors develop their own teamwork assessment tools using three instruments: (1) a team-effectiveness observation sheet, (2) a rubric for evaluating the final product, and (3) a self-assessment for students on teamwork skills	High-fidelity simulation and paper-and-pencil	• Interpersonal skills • Thinking and problem-solving skills • Product quality	Group

In other instances, models with a higher number of factors than originally hypothesized are suggested to have improved model–data fit. Aguado et al. (2014) hypothesized a five-factor model for the Teamwork Competency Test (TWCT) but obtained improved model–data fit estimates for a model containing a higher number (eight factors), which explained 56% of the total variance. The authors stated that the eight-factor solution better reproduced the analyzed data matrix compared with the five-factor model (the substantive model). The study found coverage of the content domain proposed by Stevens and Campion (1994) with Cronbach alpha > .80. The authors suggested that the TWCT yielded improved fit over the Teamwork-KSA Test, as it was able to extract a greater number of factors that were more closely aligned with the theorized model.

Predictive Validity

Results from incremental predictive validity studies revealed that measures of knowledge and skills both related to a specific task and to teamwork more generally predicted individual performance in work contexts (e.g., McClough & Rogelberg, 2003; Stevens & Campion, 1999) over and above cognitive measures. An incremental predictive validity study was conducted using the Teamwork-KSA Test with employees in real work teams using supervisor and peer ratings of job performance (McClough & Rogelberg, 2003). The Teamwork-KSA Test correlated with teamwork performance ($r = .44$, $p < .05$), with ratings of overall job performance ($r = .56$, $p < .05$), and with ratings of overall job performance ($r = .53$, $p < .05$). It also provided a significant increase in explained variance beyond aptitude measures in relation to teamwork (incremental $R^2 = .08$) and overall job performance (incremental $R^2 = .06$). The implications of this study are that the use of teamwork assessments can help augment predictions of job satisfaction and job performance over and above the use of aptitude measures alone.

Inferential Level of the Assessments

Table 4 also indicates whether the assessment was intended to make inferences about an individual's behavior within a team or the functioning of a team as a unit. In the context of developing an assessment of CPS for higher education, our focus would be on understanding and making inferences about how an individual functions within a team. As can be seen in Table 4, such assessments have previously been developed.

Overlap Between Existing Assessments and Our Proposed Definition

In Table 5, we list a sample of existing assessments of CPS skills and indicate whether the components we identified in our proposed taxonomy (see Figure 1) are measured by the existing assessments. Consistent with the idea that teamwork, collaboration, and CPS have received widely different definitions (as noted previously), each of the assessments described in Table 5 similarly evaluates various skills and behaviors under the umbrella of CPS. Note that although there is some overlap between the skills assessed by the various measures and the components listed in our proposed taxonomy, there are some components, such as active listening, that are not explicitly assessed by any of the existing measures.

Considerations in Assessing Collaborative Problem Solving

In the remainder of this section, we discuss considerations for the design of a CPS assessment. We discuss possible task types, item formats, and issues of accessibility when assessing diverse populations. This section concludes with a brief description of the possible advantages of our proposed operational definition of CPS and other assessment considerations.

Task Types

To provide authenticity, motivation, and engagement with the presented material, Grummon (1997) suggested using a variety of task types and structural features in assessment design. Dwyer, Millett, and Payne (2006) also suggested the use of a variety of assessment formats beyond the use of multiple-choice item types in alignment with the fair and valid testing of higher education skills. As a way to evaluate the possible task types that could be amenable for an assessment of CPS, we evaluated the advantages and disadvantages associated with various task types previously utilized to assess

Table 5 Existing Collaborative Problem Solving (CPS) Assessments and Their Connection to Our Proposed Taxonomy

		Taxonomy components and related skills																						
		Teamwork								Communication			Leadership						Problem solving					
Domain	Assessment	Teamwork (general)	Team cohesion	Team empowerment	Team learning	Self-management/self-leadership	Open-mindedness	Adaptability	Flexibility	Communication (general)	Active listening	Exchanging information	Leadership (general)	Organizing activities and resources	Performance monitoring	Reorganizing when faced with obstacles	Resolving conflicts	Transformational leadership	Problem solving (general)	Identifying problems	Brainstorming	Planning	Interpreting and analyzing information	Evaluating and implementing solutions
Self-assessment forced choice																								
• Military	TAPAS		√			√	√	√					√	√										
• Education and workforce	VIEW							√											√					
	WorkFORCE	√							√							√			√					
Situational judgment test (SJT)																								
• Education and workforce	TKSAT					√				√				√	√		√					√		
	PSJT	√	√			√		√																
Third-party evaluations with Likert-type items																								
• Graduate school	*ETS*® PPI	√								√			√											
Observational tools (e.g., behavioral marker systems, behavioral checklist)																								
• Health & medicine	UT BAMF	√			√			√			√	√	√											
	PACT		√								√		√	√	√		√							
	ANTS					√		√	√		√		√	√										
	CATS						√			√	√		√		√									
	OTAS					√				√		√	√	√										
	TPSAT			√		√	√										√		√	√				
• Military	ATOM					√				√	√	√		√	√									
Blended assessment																								
• Education	CCSSO		√			√											√		√					

Table 6 Advantages and Disadvantages Associated with Assessments Measuring Noncognitive Constructs

Advantages	Disadvantages
Self-assessments (Likert Type). (Brill, 2008; Callegaro, 2008; ETS, 2012; Greenleaf, 2008; OECD, 2014; Villar, 2008; Zhuang et al., 2008)	
• Convenient and economical • Compared to third-party judgments, examinees know the most about their teamwork behavior. • Can be independent from other parts of the applicants' application process.	• Response biases (social desirability, extreme, central tendency or acquiescent responding) • Fakeable • Students must have the necessary metacognitive ability to accurately gauge their own levels of teamwork
Self-assessments (forced choice). (Brown & Maydeu-Olivares, 2011; Heggestad, Morrison, Reeve, & McCloy, 2006; OECD, 2014)	
• Reduces response biases such as acquiescence or central tendency responding • Reduces "halo" effects	• Scores are ipsative rather than normative, so they cannot be compared across individuals for admissions and employment decisions
Situational judgment tests. (Cullen, Sackett, & Lievens, 2006; Lievens, Buyse, & Sackett, 2005; OECD, 2014; Whetzel & McDaniel, 2009; Zhuang et al., 2008)	
• Less adverse effects on ethnic minorities • Validity evidence supports their use in employment and educational settings. • Direct measure of judgment making • Possibly more engaging than their self-report style counterparts	• They may be easy to fake if they contain behavioral tendency instructions instead of knowledge instructions. • Test takers are able to slightly increase their score through retesting or coaching
Third-party evaluations. (Kyllonen, 2008; Zhuang et al., 2008)	
• Less expensive • Scalable	• Prone to response biases (e.g., halo effects) • Responses are not comparable between instructors and students
Observational tools. (Daimler-Und & Benz-Stiftung, 2001)	
• Direct measure of teamwork performance in context • Fake resistant	• Raters need to be recruited, trained, calibrated, and periodically recalibrated • Cannot capture every aspect of performance because not all behaviors may be elicited during observation, or the situations may be too complex to observe every behavior • Difficult to standardize the test conditions • Very expensive for large-scale administrations

various noncognitive skills. As summarized in Table 6, these task types are (a) self-assessments with Likert scales; (b) self-assessments with forced-choice, (c) situational judgment tests (SJTs); (d) third-party evaluations; and (e) observational tools. The table also lists our evaluation of the task types with respect to whether the assessments are (a) resistant to faking, gaming, and coaching; (b) scalable; (c) cost-efficient to produce and score; (d) resistant to the various biases that have been documented in the research; and (e) secure.

Upon our evaluation, two task types (third-party evaluations and observational tools) received a less favorable evaluation on the abovementioned criteria. First, although the advantage to third-party evaluations is the removal of the

individual as the evaluator, these task types have other limitations. One is the difficulty of controlling response patterns or biases possibly held by individual instructors. Moreover, low reliability has been reported when comparing instructors' ratings, as the evaluations are possibly based on subjective judgments (Zhuang, MacCann, Wang, Liu, & Roberts, 2008). Third, halo effects may occur if an instructor has a positive impression about one or more of an examinee's attributes (e.g., creativity), which generalize across all other attributes (Nisbett & Wilson, 1977).

Observation tools also received low ratings as a possible contender due to challenges related to scoring expense and the variability of testing conditions. As O'Neil et al. (1997) indicated, these methods are neither practical nor cost effective in large-scale test settings. They can be expensive given the need to recruit, train, and calibrate raters (Daimler-Und & Benz-Stiftung, 2001). This is likely to outweigh the benefits of reduced fakeability and their ability to yield direct measurements of teamwork performance. A second limitation is that not all behaviors may be displayed during the assessment time frame. For example, it would be difficult to assess a test taker's conflict resolution skills if team members are largely cooperative throughout the exercise. In the same vein, a third limitation is the difficulty of standardizing the test conditions because an examinee's performance might be impacted by his or her interactions with teammates or their teammates' behavior. In the context of admissions or any other administration with large numbers of subjects, the use of observation tools would be expensive and difficult to operationalize. Such tools would require that multiple examinees interact with each other in controlled situations (in an attempt to standardize the required tasks), and there are no assurances that the behaviors of interest would be elicited for evaluation. Additional limitations include the need for transcribing, coding, and analyzing information post hoc, precluding the analysis of information expediently and thus delaying the reporting of an individual's performance within a team.

Self-Assessments Using Forced Choice

Self-assessments using Likert scales are one of the most commonly used task types to assess noncognitive skills, given their low cost and convenient administration (O'Neil et al., 2004; Zhuang et al., 2008). However, they are easily fakeable and coachable, and they depend on "students' capabilities for self-knowledge: Students must have the necessary psychological-mindedness to accurately gauge their own levels of teamwork" (Zhuang et al., 2008, p. 5). These types of assessments are also prone to response biases that could occur in terms of an examinee's interpretation or use of the scale, particularly in cross-cultural contexts. Other response biases may occur because of social desirability, tendencies toward endorsing a particular part of the scale (extreme or central tendencies), or acquiescent response patterns, where the test taker endorses a particular item due to social desirability (ETS, 2012; OECD, 2012). As an alternative, the use of self-assessments using forced choice might be preferable to those using Likert scales, as they may be harder to game, and they are relatively inexpensive to produce and quick to score. They can help reduce response biases such as tendencies toward acquiescence or central response tendencies because this format makes it impossible to endorse every item in the same way (Cheung & Chan, 2002).

As an example, a forced-choice item may contain the following two statements, from which the examinee must choose one statement over the other: (a) I can relax easily, or (b) I set high personal standards for myself. The selection of one of the two responses can also help reduce the effects of the use of high ratings across domains because the option choices are not attributes or traits at opposite ends of a scale but instead are on traits that are on different dimensions (Bartram, 2007; Brown & Maydeu-Olivares, 2011, 2013; OECD, 2012). Previous research has suggested that they are more fake resistant, as one among two choices needs to be selected. Both choices may appear equally (un)appealing (Bowen, Martin, & Hunt, 2002; Brown & Maydeu-Olivares, 2013; Christiansen, Burns, & Montgomery, 2005; Jackson, Wroblewski, & Ashton, 2000; OECD, 2012; White & Young, 1998). Beyond these advantages, their inclusion within a blended assessment approach would be helpful by increasing the amount of data collected as a cross-check for other sections of the assessment. It could also help increase assessment reliability through the administration of additional items.

A downside is that they are not based on direct observations of an examinee's behaviors. They typically do not have normative scores, as scores in one dimension are relative only to scores on different dimensions for the same individual and not to other examinees (Brown & Maydeu-Olivares, 2011). This downside limits the ability to compare individuals, which is important in higher education admissions (Heggestad et al., 2006), further suggesting the need to complement this task type with others. Current research might help remedy this shortcoming. See the following approaches (a) Stark, Chernyshenko, and Drasgow's multi-unidimensional pairwise-preference model (2005); (b) de la Torre, Ponsoda, Leenen, and Hontangas' (2011) extension of Stark et al.'s model; and (a) Brown and Maydeu-Olivares (2013).

Situational Judgment Tests

SJTs composed of short vignettes have also been used widely in multiple fields, such as business and education. Their advantages include their usefulness in detecting subtle judgment processes by asking participants to provide intuitive or contextual judgments about scenarios set in plausible contexts. This contextualization in real-life scenarios also renders them more engaging than other task types, such as self-reports (Zhuang et al., 2008). Weekly and Ployhart (2005) suggested they can enhance the incremental predictive validity of traditional personality and cognitive measures.

SJTs are flexible and may use a variety of response formats and delivery modes that extend beyond the use of multiple-choice and render them more engaging. Possible response formats include the use of constructed-response items or chat boxes that present challenging and possibly complex situations to probe into the examinee's ability to display the behaviors and skills thought central to CPS. SJTs can be delivered online, which increases the possibility to elicit behaviors as examinees interact with avatars or simulated team members, thereby increasing the authenticity of the assessment and performances. They can also be more dynamic and enable the use of various pathways based on an examinee's responses.

Despite the advantages, SJTs also have three main limitations, which might present challenges in their use in assessing CPS: (a) fakeability, (b) coachability, and (c) item authoring. First, although SJTs with behavior tendency items have higher correlations with personality than items eliciting knowledge, they may be easier to fake (Nguyen, Biderman, & McDaniel, 2005). Second, in a high-stakes setting, Lievens, Buyse, Sackett, and Connelly (2012) found an incremental effect ($SD = 0.5$) between coaching and self-test preparation across alternate forms of an interpersonal SJT in the context of a medical school admissions test. Third, challenges arise in relation to how to define someone as an expert in teamwork for the item authoring and scoring, as different situations may lend themselves to different kinds of expertise; hence, finding item writers to create a diverse range of plausible scenarios might be difficult (Zhuang et al., 2008).

Simulated Scenario-Based Tasks

This task type provides the opportunity to observe examinees' behaviors in a presented scenario. This approach is closer to direct observation and can potentially yield data such as desirable or undesirable behaviors. Second, simulated scenario-based tasks may be more engaging and realistic for the examinees. Third, simulated scenarios allow for the control of the situation to the degree that the presented situation is the same for all examinees. Each scenario can have decision points embedded within it in order to track and score an examinee's choices. Chat screens or avatars can be embedded to gauge how examinees interact with simulated teammates (Y. Rosen & Tager, 2013).

Although simulated scenario-based tasks offer increased test-taker engagement, authenticity, and standardization over other item types, the possibility of "gaming" the task persists. Examinees may respond to the test based on guesses of what the desired responses or outcomes are rather than what they would do under regular conditions. For example, a very bright and intuitive examinee might "know" what the desired attributes of someone with high collaborative skills are and guess what the "correct" pathway is for the task set, even if the student was reluctant to collaborate with peers and mentors in nonassessment settings. For this reason, it might not be advisable to use this task type in isolation. A second consideration is the cost associated with developing and implementing the scenarios, which on one hand are advantageous for rendering pseudoauthentic responses, but on the other hand may be difficult to keep secure.

Recommended Approach to Collaborative Problem Solving Assessment

In light of our review, we suggest using a blended assessment approach to assess CPS in (high-stakes) large-scale assessment contexts. It might consist of various combinations of task types, such as (a) forced-choice self-assessment, (b) SJTs composed of vignettes, and (c) a simulated scenario-based task. These were our top three task-type contenders. The blended approach can be beneficial in relation to accessing the strengths of the different task types while balancing their weaknesses. The Council of Chief State School Officers (CCSSO) Workplace Readiness Assessment Consortium used a blended approach previously (Grummon, 1997).

Accessibility and the Reduction of Sources of Construct-Irrelevant Variance

Beyond the selection of task types and structural features to ensure high levels of examinee motivation and engagement with the task, it is also important from a fairness and validity standpoint that the items are accessible to the diverse

populations taking the assessment, such as students with disabilities or students who are culturally and linguistically diverse. Guidelines exist for the development of assessments that are sensitive to students with learning disabilities, which should be consulted in the development of prototypes of an assessment of CPS (Davey, 2011; Stone & Davey, 2011). Taking into account students with disabilities is of utmost importance for fairness purposes, particularly given that these students are attending higher education institutions at increasingly elevated rates, as suggested by Heiman and Precel (2003).

Several guidelines have been developed to provide guidance on the development of fair and valid assessments for diverse test-taker populations. For instance, the International Test Commission (2010) included two particularly relevant guidelines: C1 and D1. Respectively, they state, "Effects of cultural differences which are not relevant or important to the main purposes of the study should be minimized to the extent possible," and "Test developers/publishers should ensure that the adaptation process takes full account of linguistic and cultural differences among the populations for whom adapted versions of the test or instrument are intended" (p. 2). Such guidelines should be reviewed in the development of the prototypes of assessments of CPS for higher education. Sensitivity to cultural and linguistic differences is particularly important in the context of assessing higher education students, given the large number of international populations taking these kinds of assessments. For instance, in the 2015–2016 administration of the *GRE*® General Test, 43% of the population comprised non-U.S. citizens, an increase of 15% from the 2011–2012 administration (ETS, 2014a, 2016).

Previous research has also provided guidance on how to reduce construct-irrelevant variance due to cultural differences in assessment development. For example, PISA used various strategies to reduce cultural load when designing item features and content. These issues were of central importance in PISA, as it is administered to multiple countries globally. Consideration was made in the way items were written and selected as well as in the development of PISA's conceptual framework, which we used to inform our proposed operational definition. Such efforts are helpful in informing the development of task types and design patterns (Oliveri, Lawless, & Mislevy, 2017).

Concluding Note

In this paper, we presented a synthesis of the CPS literature and provided our perspectives on the components, skills, and behaviors that are related to CPS within the context of higher education readiness, with connections to the workforce. We provided an operational definition of CPS and discussed assessment considerations to capture the behaviors and elements of CPS that are generalizable across tasks and teammates. Accordingly, our literature review focused on identifying and describing such skills in alignment with the use of a test for higher education admissions across fields. Because of the multiple fields and subjects studied in higher education, we suggested focusing on taxonomies such as Aguado et al. (2014), Cannon-Bowers et al. (1995), and Stevens and Campion (1994, 1999), who discussed taxonomies of transportable skills. A focus on transportable skills is important, as meaningful and practical gains in workforce readiness of college students were found to relate to gains in CPS. As stated by Chen, Donahue, and Klimoski (2004), a systematic focus on teaching and learning of CPS in a wider number of university curricula across fields such as business, engineering, and health care could potentially improve workforce readiness of college graduates, which may in turn translate into better teamwork behavior in actual work settings. Such efforts could be useful in assisting universities in their efforts to prepare students for today's and the future workplace.

Although we envision rigorous work will need to be done to ensure meeting the psychometric standards of the assessment, we suggest that a first step of this effort is to provide conceptual clarity—in other words, an operational definition that provides sufficiently concrete terms to help inform measurement efforts. Exemplars of existing assessments do not delineate constructs or behaviors with sufficient clarity to develop substantive models. We noted this absence particularly in relation to higher education. We thus aimed to provide an operational definition that would provide sufficient information to lay out the foundation to develop a substantive model that later could be empirically assessed. In so doing, this paper supports the efforts of the scholars who have begun conceptual work intended to clarify the nature and the dimensional structure of teamwork processes (e.g., Marks et al., 2001). Such research (although currently sparse) is central to advancing the measurement and assessment of CPS to support construct validity. We also suggest the use of a blended approach to assessing CPS. This approach is not new, and the task types (e.g., forced-choice self-assessments, SJTs, and scenario-based) we wish to use have already been developed to measure other constructs across various contexts.

Acknowledgments

The authors would like to thank Vinetha Belur, Kri Burkander, Chelsea Ezzo, and Laura Ridolfi for their help with the literature searches and research support, and Robert Mislevy, Patrick Kyllonen, Alina von Davier, Katrina Roohr, Lei Lu, Paul Borysewicz, Peter Cooper, and Eric Steinhauer for their consultation.

Note

1 Henceforth, all terminology and skills used interchangeably in the literature with *teamwork* or *CPS* will be referred to simply as CPS.

References

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2014). *The Degree Qualifications Profile: A learning-centered framework for what college graduates should know and be able to do to earn the associate, bachelor's, or master's degree*. Indianapolis, IN: Lumina Foundation.

Aguado, D., Rico, R., Sánchez-Manzanares, M., & Salas, E. (2014). Teamwork Competency Test (TWCT): A step forward on measuring teamwork competencies. *Group Dynamics: Theory, Research, and Practice*, *18*(2), 101 – 121.

Association of American Colleges and Universities. (2011). *The LEAP vision for learning: Outcomes, practices, impact, and employers' views*. Washington, DC: Author.

Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment, 15,* 263 – 272.

Bowen, C. C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis*, *10*(3), 240 – 259.

Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1993, April). *Impact of raters and events on team performance measurement.* Paper presented at the eighth annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Brill, J. E. (2008). Likert Scale. In Paul J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 427 – 429). Thousand Oaks, CA: Sage.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*(3), 460 – 502. doi: 10.1177/0013164410375112

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, *18*(1), 36 – 52.

Burrus, J., Jackson, T., Xi, N., & Steinberg, J. (2013). *Identifying the most important 21st century workforce competencies: An analysis of the Occupational Information Network (O*NET)* (Research Report No. RR-13-21). Princeton, NJ: Educational Testing Service. 10.1002/j.2333-8504.2013.tb02328.x

Callegaro, M. (2008). Social desirability. In Paul J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 825 – 826). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412963947.n486

Cannon-Bowers, J. A., & Salas, E. (1997). A framework for developing team performance measures in training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 45 – 62). Mahwah, NJ: Lawrence Erlbaum.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. *Team Effectiveness and Decision Making in Organizations*, *333*, 380.

Casner-Lotto, J., & Barrington, M. (2006). *Are they really ready for work? Employers' perspectives on the basic knowledge and applied skills of new entrants into the 21st century workforce*. New York, NY: Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families, and Society for Human Resource Management.

Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training undergraduates to work in organizational teams. *Academy of Management Learning and Education, 3*, 27 – 40.

Cheung, M. W. L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling, 9*, 55 – 77.

Chiu, C.-J., Brock, D., Abu-Rish, E., Vorvick, L., Wilson, S., Hammer, D., … , Zierler, B. (2011). *Performance Assessment of Communication and Teamwork (PACT) tool set*. Retrieved from http://collaborate.uw.edu/educators-toolkit/tools-for-evaluation/performance-assessment-of-communication-and-teamwork-pact-too

Chiu, M. M. (2000). Group problem-solving processes: Social interactions and individual actions. *Journal for the Theory of Social Behaviour, 30*(1), 26 – 49. doi:10.1111/1468-5914.00118

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering the use of forced-choice formats for applicant personality assessment. *Human Performance*, *18*, 267 – 307.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human Factors, 42*(1), 151 – 173.

Council of Graduate Schools. (2012). *U.S. must close gap between graduate schools, employers to stay competitive, spur innovation* [Press release]. Retrieved from http://www.cgsnet.org/pathways_release

Creative Problem Solving Group, Inc. (2013). *About VIEW: An assessment of problem solving style*. Retrieved from http://www.viewassessment.com/media/978/About-VIEW-2014.pdf

Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, *14*(2), 142 – 155.

Daimler-Und, G., & Benz-Stiftung, K. (2001, July 5 – 6). *Enhancing performance in high risk environments: Recommendations for the use of behavioural markers* [Workshop]. Zurich, Switzerland: SwissAir Training Centre.

Davey, T. (2011). *Practical considerations in computer-based testing*. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/CBT-2011.pdf

de la Torre, J., Ponsoda, V., Leenen, I., & Hontangas, P. (2011, July). *Some extensions of the multi-unidimensional pairwise-preference model.* Paper presented at the 26th Annual Conference for the Society of Industrial and Organizational Psychology, Chicago, IL.

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 19 – 43). Mahwah, NJ: Lawrence Erlbaum.

Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C. L., & White, L. A. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions* (Technical Report No. 1311). Fort Belvoir, VA: Army Research Institute for the Behavioral and Social Sciences.

Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes.* Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2009). *About the ETS® Personal Potential Index (ETS® PPI)*. Retrieved from http://www.ets.org/ppi/evaluators/about/

Educational Testing Service. (2012). *Assessment methods*. Retrieved from http://www.ets.org/s/workforce_readiness/pdf/21333_big_5.pdf

Educational Testing Service. (2014a). *A snapshot of the individuals who took the GRE® revised General Test*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2014.pdf

Educational Testing Service. (2014b, Nov.). Innovative Assessment of Collaboration 2014 Meeting Program. *Innovative Assessment of Collaboration Conference 2014*. Arlington, VA: Educational Testing Service. Retrieved from https://custom.cvent.com/DE290BCAC7FE4C9D8E112A32B558BCC1/files/2029ac56dba04f1ca0fccb03b8185410.pdf

Educational Testing Service. (2016). *A snapshot of the individuals who took the GRE® revised General Test*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2016.pdf

Ennis, M. R. (2008). *Competency models: A review of the literature and the role of the Employment and Training Administration (ETA)*. Washington, DC: U.S. Department of Labor.

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 52*(2), 203 – 224.

Flin, R., Glavin, R., Maran, N., & Patey, R. (2012). *Framework for observing and rating anaesthetists' non-technical skills: Anaesthetists' non-technical skills (ANTS) system handbook, v 1.0*. Retrieved from http://www.abdn.ac.uk/iprc/documents/ANTS%20Handbook%202012.pdf

Frankel, A., Gardner, R., Maynard, L., & Kelly, A. (2007). Using the Communication and Teamwork Skills (CATS) assessment to measure health care team performance. *Joint Commission Journal on Quality and Patient Safety/Joint Commission Resources*, *33*(9), 549 – 558.

González, J., & Wagenaar, R. (Eds.). (2003). *Tuning educational structures in Europe. Final Report Pilot Project - Phase 1*. Retrieved from http://tuningacademy.org/wp-content/uploads/2014/02/TuningEUI_Final-Report_EN.pdf

Greenleaf, E. A. (2008). Extreme response style. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (p. 257). Thousand Oaks, CA: Sage. 10.4135/9781412963947.n486

Grummon, P. T. H. (1997). Assessing teamwork skills for workforce readiness. In H. J. O'Neil (Ed.), *Workforce readiness: Competencies and assessment* (pp. 383 – 410). Mahwah, NJ: Lawrence Erlbaum.

Hart Research Associates. (2015). *Falling short? College learning and career success: Selected findings from online surveys of employers and college students conducted on behalf of the Association of American Colleges & Universities*. Retrieved from https://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*, 9–24.

Heiman, T., & Precel, K. (2003). Students with learning disabilities in higher education: Academic strategies profile. *Journal of Learning Disabilities*, *36*(3), 248–258.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach*. Dordrecht, The Netherlands: Springer.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

International Society for Technology in Education. (n.d.). *ISTE standards. Students*. Retrieved from http://www.iste.org/docs/pdfs/20-14_ISTE_Standards-S_PDF.pdf

International Test Commission. (2010). *International Test Commission guidelines for translating and adapting tests*. Retrieved from https://www.intestcom.org/files/guideline_test_adaptation.pdf

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*(4), 371–388.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403–437.

Knapp, D. J., & Heffner, T. S. (Eds.) (2009). *Validating future force performance measures (Army class): End of training longitudinal validation* (Technical Report No. 1257). Arlington, VA: Army Research Institute for the Behavioral and Social Sciences.

Kostopoulos, K. C., Spanos, Y. E., & Prastacos, G. P. (2013). Structure and function of team learning emergence: A multilevel empirical validation. *Journal of Management, 39*(6), 1430–1461.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, *7*, 77–124.

Kukenberger, M. R., Mathieu, J. E., & Ruddy, T. (2015). A cross-level test of empowerment and process influences on members' informal learning and team commitment. *Journal of Management*, *46*, 227–259.

Kyllonen, P. C. (2008). *The research behind the ETS® Personal Potential Index (PPI)*. Retrieved from http://www.ets.org/Media/Products/PPI/10411_PPI_bkgrd_report_RD4.pdf

Kyllonen, P. C. (2012, May). *Measurement of 21st century skills within the common core state standards*. Paper presented at the Invitational Research Symposium on Technology Enhanced Assessments. Princeton, NJ. Retrieved from https://www.ets.org/Media/Research/pdf/session5-kyllonen-paper-tea2012.pdf

Lench, S., Fukuda, E., & Anderson, R. (2015). *Essential skills and dispositions: Developmental frameworks for collaboration, creativity, communication, and self-direction*. Lexington: Center for Innovation in Education at the University of Kentucky.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, *58*(4), 981–1007. doi:10.1111/j.1744-6570.2005.00713.x

Lievens, F., Buyse, T., Sackett, P. R., & Connelly, B. S. (2012). The effects of coaching on situational judgment tests in high-stakes selection. *International Journal of Selection and Assessment*, *20*, 272–282. doi:10.1111/j.1468-2389.2012.00599.x

Markle, R., Brenneman, M., Jackson, T., Burrus, J., & Robbins, S. (2013). *Synthesizing frameworks of higher education student learning outcomes* (Research Report No. RR-13-22). Princeton, NJ: Educational Testing Service. 10.1002/j.2333-8504.2013.tb02329.x

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356–376.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 272–283.

McClough, A. C., & Rogelberg, S. G. (2003). Selection in teams: An exploration of the teamwork knowledge, skills, and ability test. *International Journal of Selection and Assessment*, *11*(1), 56–66. doi:10.1111/1468-2389.00226

Mitchell, R., Boyle, B., Parker, V., Giles, M., Joyce, P., & Chiang, V. (2014). Transformation through tension: The moderating impact of negative affect on transformational leadership in teams. *Human Relations, 67*(9), 1095–1121. doi:10.1177/0018726714521645

Murray, T. S., Owen, E., & McGaw, B. (2005). *Learning a living: First results of the adult literacy and life skills survey*. Ottawa, Canada: Statistics Canada and the Organisation for Economic Co-operation and Development.

National Research Council. (2010). *Assessing 21st century skills: Summary of a workshop.* J. A. Koenig, Rapporteur, Committee on the Assessment of 21st Century Skills, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment, 13*, 250–260.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*, 250–256.

Oliveri, M. E., Lawless, R. R., & Mislevy, R. J. (2017). *Using evidence-centered design to support the development of culturally and linguistically sensitive collaborative problem-solving assessments.* Manuscript in preparation.

Oliveri, M. E., McCaffrey, D., Holtzman, S., & Ezzo, C. (2014, April). *Investigating the factor structure of the Personal Potential Index using a multilevel factor analysis approach. In M. E. Oliveri (Chair), Problems with interpretations of multilevel data – Extending research beyond hierarchical linear modeling.* Paper presented at the annual American Educational Research Association conference, Philadelphia, PA.

O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2004). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education, 10*, 361 – 373.

O'Neil, H. F., Chung, G., & Brown, R. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411 – 452). Mahwah, NJ: Lawrence Erlbaum.

Organisation for Economic Co-operation and Development. (2012). *Literacy, numeracy and problem solving in technology-rich environment. Framework for the OECD Survey of Adult Skills.* http://dx.doi.org/10.1787/9789264128859-en

Organisation for Economic Co-operation and Development. (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy.* https://doi.org/10.1787/9789264190511-en

Organisation for Economic Co-operation and Development. (2013b). *PISA 2015. Draft collaborative problem solving framework*. Retrieved from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Organisation for Economic Co-operation and Development. (2014). Context questionnaire development. In R. Turner (Ed.), *PISA 2012 Technical Report* (pp. 47 – 64). Retrieved from http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

Oser, R. L., McCallum, G. A., Salas, E., & Morgan, B. Jr. (1989). *Toward a definition of teamwork: An analysis of critical team behavior* (NTSC Technical Report No. 90-009). Orlando, FL: Naval Training System Centre.

Partnership for 21st Century Skills. (2012). *Framework for 21st century learning.* Retrieved from http://www.p21.org/our-work/p21-framework

Rosen, M. A., Weaver, S. J., Lazzara, E. H., Salas, E., Wu, T., Silvestri, S., … , King, H. B. (2010). Tools for evaluating team performance in simulation-based training. *Journal of Emergencies, Trauma & Shock*, *3*(4), 353 – 359. doi:10.4103/0974-2700.70746

Rosen, Y., & Tager, M. (2013). *Computer-based assessment of collaborative problem-solving skills: Human-to-agent versus human-to-human approach*. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/collaborativeproblemsolvingresearchreport.pdf

Rotondi, A. J. (1999). Assessing a team's problem solving ability: Evaluation of the Team Problem Solving Assessment Tool (TPSAT). *Health Care Management Science, 2*, 205 – 214.

Rousseau, V., & Aubé, C. (2010). Team self-managing behaviors and team effectiveness: The moderating effect of task routineness. *Group & Organization Management, 35*(6), 751 – 781.

Salisbury, W. D., Carte, T. A., & Chidambaram, L. (2006). Cohesion in virtual teams: Validating the perceived cohesion scale in a distributed setting. *ACM SIGMIS Database*, *37*(2 – 3), 147 – 155.

Shanahan, C., Best, C., Finch, M., & Sutton, C. (2007). *Measurement of the behavioural, cognitive, and motivational factors underlying team performance* (DSTO Research Report DSTO-RR-0328). Fishermans Bend Victoria, Australia: Air Operations Division Defence Science and Technology Organisation.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 61 – 87). Washington, DC: APA Press.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, *29*(3), 184 – 203.

Stevens, M. J., & Campion, M. A. (1994). The knowledge, skills and ability requirements for teamwork: Implications for human resources management. *Journal of Management, 20*(2), 502 – 528.

Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, *25*(2), 207 – 228.

Stone, E., & Davey, T. (2011). Computer-adaptive testing for students with disabilities: A review of the literature (RR-11-32). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2011.tb02268.x

Strayhorn, T. L. (2006). *Frameworks for assessing learning and development outcomes*. Washington, DC: Council for the Advancement of Standards in Higher Education.

Thomas, E. J., Sexton, J. B., & Helmreich, R. L. (2004). Translating teamwork behaviours from aviation to healthcare: Development of behavioural markers for neonatal resuscitation. *Quality Safety Health Care, 13*(1), i57-i64. doi:10.1136/qshc.2004.009811

Treffinger, D. J., Isaksen, S. G., & Selby, E. C. (2014). *Evidence supporting VIEW*. Orchard Park, NY: Creative Problem Solving Group.

Villar, A. (2008). Response bias. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 752 – 754). Thousand Oaks, CA: Sage. doi:10.4135/9781412963947.n486

Waugh, G. W., & Russell, T. L. (2005). Predictor situational judgment test. In D. J. Knapp & T. R. Tremble (Eds.) *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Weekly, J. A., & Ployhart, R. E. (Eds.). (2005). *Situational judgment tests: Theory, measurement, and application*. Hillsdale, NJ: Lawrence Erlbaum.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188–202.

White, L. A., & Young, M. C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM)*. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330,* 686–688. doi:10.1126/science.1193147

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *Leadership Quarterly, 12*, 451–483.

Zhuang, X., MacCann, C., Wang, L., Liu, O. L., & Roberts, R. D. (2008). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students* (Research Report RR-08-50). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2008.tb02136.x

Suggested citation:

Oliveri, M. E., Lawless, R., Molloy, H. (2017). *A literature review on collaborative problem solving for college and workforce readiness* (GRE Board Report No. 17-03). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12133

Action Editor: Brent Bridgeman

Reviewers: This report was reviewed by the GRE Technical Advisory Committee and the Research Committee and Diversity, Equity and Inclusion Committee of the GRE Board.